STOR 565 Homework

1. (Basic properties of the inner product) Let $x, y, z \in \mathbb{R}^d$ and $a, b \in \mathbb{R}$. Show that

   a. $\langle x, y \rangle = \langle y, x \rangle$

   b. $\langle ax, by \rangle = ab \langle x, y \rangle$

   c. $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$

   d. $||x + y||^2 = ||x||^2 + 2\langle x, y \rangle + ||y||^2$

2. (Data and sample covariance matrices) Let $\mathbf{S}$ be the sample covariance matrix of a data set $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Answer/verify the following. You may repeat the arguments given in class, but clearly explain your work.

   a. Define the data matrix $\mathbf{X}$

   b. Give the definition of $\mathbf{S}$ in terms of the data matrix $\mathbf{X}$

   c. Show that $\mathbf{S} = n^{-1} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^t$

   d. $\mathbf{S}$ is symmetric and non-negative definite

   e. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ be the eigenvalues of $\mathbf{S}$. Show that $\sum_{k=1}^{p} \lambda_k = n^{-1}||\mathbf{X}||^2$

   f. $\text{rank}(\mathbf{S}) = \text{rank}(\mathbf{X}^t\,\mathbf{X}) = \text{rank}(\mathbf{X}) \leq \min(n, p)$

   g. If $p > n$ then $\text{rank}(\mathbf{S}) < p$ and $\mathbf{S}$ is not invertible.

3. Let $\mathbf{u}_1 = (-1, 2, 0)$ and $\mathbf{u}_2 = (2, 4, 3)$. Find the projections of $\mathbf{u}_1$ and $\mathbf{u}_2$ onto $\mathbf{v}$ where:

   a. $\mathbf{v}$ = (0,1,0)

   b. $\mathbf{v}$ = (1,1,1)

   c. $\mathbf{v}$ = (1,0,-1)

4. Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^d$ be orthonormal vectors with span $V = \{\alpha\mathbf{v}_1 + \beta\mathbf{v}_2 : \alpha, \beta \in \mathbb{R}\}$. For $\mathbf{u} \in \mathbb{R}^d$ define the projection of $\mathbf{u}$ onto V to be the vector $v \in V$ that is closest to $\mathbf{u}$,

$$\text{proj}_V(\mathbf{u}) = \underset{v \in V}{\text{argmin}}\, ||u - v||.$$

Show that $\text{proj}_V(\mathbf{u}) = \langle \mathbf{u}, \mathbf{v}_1 \rangle \, \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle \, \mathbf{v}_2$. Hint: Adapt the argument used in class for the projection onto a one-dimensional subspace.

5. Consider a data set consisting of four points in $\mathbb{R}^2$

$$\mathbf{x}_1^t = (1, 2), \; \mathbf{x}_2^t = (-1, 2), \; \mathbf{x}_3^t = (2, -1), \; \mathbf{x}_4^t = (2, 1)$$

   a. Write down the data matrix $\mathbf{X}_0$ associated with this data set

   b. Column center $\mathbf{X}_0$ so that each column has mean zero. This is equivalent to replacing each observation $\mathbf{x}_i$ by the centered observation $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \frac{1}{4}\sum_{i=1}^{4} \mathbf{x}_i$. Check that $\sum_{i=1}^{4} \tilde{\mathbf{x}}_i = \mathbf{0}$, and draw a plot of the points $\tilde{\mathbf{x}}_i$. Call the recentered data matrix $\mathbf{X}$.

   c. Calculate the sample covariance matrix $\mathbf{S} = \frac{1}{4}\mathbf{X}^T \mathbf{X}$.

   d. Calculate the eigenvalues of $\mathbf{S}$. Is $\mathbf{S}$ invertible? If so, find $\mathbf{S}^{-1}$.

   e. Find orthonormal eigenvectors of $\mathbf{S}$.

   f. What is the best one-dimensional subspace (line) for approximating the centered observations $\tilde{\mathbf{x}}_i$? Draw this line on your plot.

6. Let $\mathbf{A}$ be a 3×3 lower triangular matrix given below:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 2 & 0 \\ 2 & 4 & 3 \end{pmatrix}$$

   a. What are the eigenvalues of $\mathbf{A}$?

   b. Find $\det(\mathbf{A})$?

   c. Argue that $\mathbf{A}^{-1}$ exists and calculate it.

7. (Norms of outer products) Find an expression relating the norm of the outer product $||\mathbf{u}\mathbf{v}^t||$ to the norms of the vectors $||\mathbf{u}||$ and $||\mathbf{v}||$.